Applying Functional Genomics to Chronic Obstructive Pulmonary Disease

Edwin K. Silverman

Channing Division of Network Medicine and Pulmonary and Critical Care Medicine Division, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts

Abstract

Genome-wide association studies (GWAS) have identified more than 20 genomic regions associated with chronic obstructive pulmonary disease (COPD) susceptibility. However, the functional genetic variants within these COPD GWAS loci remain largely unidentified, thus limiting translation of these GWAS discoveries to new disease insights. Whole-exome and whole-genome sequencing studies have the potential to identify rare genetic determinants of COPD. Efforts to understand the biological effects of novel COPD

genetic loci include gene-targeted murine models, integration of additional omics data (including transcriptomics and epigenetics), and functional variant identification. COPD genetic determinants likely act through biological networks, and a variety of network-based approaches have been used to gain insights into COPD susceptibility and heterogeneity.

Keywords: chronic obstructive pulmonary disease; genetics; association analysis; integrative genomics; network medicine

(Received in original form August 8, 2018; accepted in final form August 21, 2018)

Supported by National Institutes of Health grants U01 HL089856, R01 HL113264, P01 HL114501, R33 HL120794, R01 HL133135, and R01 HL137927.

Correspondence and requests for reprints should be addressed to Edwin K. Silverman, M.D., Ph.D., Channing Division of Network Medicine and Pulmonary and Critical Care Medicine Division, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, 181 Longwood Avenue, Boston, MA 02115. E-mail: ed.silverman@channing.harvard.edu.

Ann Am Thorac Soc Vol 15, Supplement 4, pp S239–S242, Dec 2018

Copyright © 2018 by the American Thoracic Society

DOI: 10.1513/AnnalsATS.201808-530MG

Internet address: www.atsjournals.org

Although chronic obstructive pulmonary disease (COPD) is strongly influenced by cigarette smoking, several lines of evidence suggest that genetic susceptibility predisposes to COPD risk. Among smokers, COPD clusters in families (1). Alpha-1 antitrypsin (AAT) deficiency and cutis laxa are monogenic syndromes that often include emphysema as part of their syndrome constellation. Over the past 30 years, COPD genetic studies have evolved from linkage analysis to candidate gene association studies to genome-wide assessments. We review the current evidence for specific genetic determinants of COPD as well as functional studies to identify key genes and functional variants related to COPD susceptibility.

Genome-Wide Association Studies of COPD and Lung Function Levels

Genome-wide association studies (GWAS), which test commercialized panels of hundreds of thousands of single-nucleotide polymorphisms (SNPs) in large samples of cases/control subjects, population cohort members, or family units, have been successfully applied in COPD (2). With the large number of statistical tests performed in GWAS, stringent adjustment for multiple statistical testing is required; P values less than 5×10^{-8} are typically designated as genome-wide significant. A recent report from the International COPD Genetics Consortium identified 22 genome-wide significant regions associated with

spirometrically defined COPD (3). Of note, nine of these regions had been previously associated with COPD in GWAS, and nine other regions had been previously associated with lung function levels but not COPD. It has been reassuring that a number of these COPD GWAS loci have been replicated in additional study populations by other investigators, suggesting that these associations are valid. In parallel with these GWAS of COPD cases and control subjects, large general population GWAS of spirometric measures have identified an even larger number of genetic determinants of forced expiratory volume in 1 second (FEV₁) and FEV₁/forced vital capacity (4). The significant overlap between COPD and lung function determinants has been both unsurprising—COPD was defined by lung

function, after all—and remarkable, because it is not apparent how genetic determinants of small effect on lung function could lead to the profound reductions in lung function and pathological abnormalities (e.g., emphysema, small airway destruction and fibrosis) seen in advanced COPD. Although the impact of individual COPD GWAS determinants on COPD risk is modest, genetic risk scores hold promise for providing a more profound predictive capacity for the presence and severity of COPD (5).

Rare Genetic Determinants of COPD

With the substantial insights into COPD pathobiology resulting from AAT deficiency, other rare genetic determinants of COPD were searched for using whole-exome and whole-genome sequencing (WGS) studies. Stanley and colleagues identified rare variants in the *TERT* gene, previously associated with idiopathic pulmonary fibrosis, to be associated with emphysema (6). Of interest, their functional studies of specific genetic variants were essential; without the ability to separate three nonsynonymous variants that functionally influenced telomere length in COPD cases from one nonsynonymous variant in control subjects that did not impact telomere length, this relationship would likely not have been identified. Similar detailed functional insights of specific genetic variants are rarely available for most genes implicated through DNA sequencing studies.

Qiao and colleagues analyzed whole-exome sequencing in extended pedigrees ascertained through severe, early-onset COPD cases without AAT deficiency (7).

They found 69 genes that had potentially functional variants segregating with COPD in at least two families but no genes that had variants segregating in more than three families. These results suggested that rare coding genetic determinants of COPD are likely quite genetically heterogeneous.

Large-scale WGS analyses of COPD are currently being performed through the U.S. National Heart, Lung, and Blood Institute TOPMed (Trans-Omics in Precision Medicine) program. These WGS efforts are unlikely to find many more novel common variant associations for COPD and COPD-related phenotypes than GWAS, except possibly in African American subjects, because the widely used genome-wide SNP genotyping panels for GWAS provide less-adequate coverage for populations of African ancestry. These WGS efforts will be very helpful for statistical fine mapping of previously identified GWAS regions for COPD and COPD-related phenotypes (8), and they will also enable identification of rare coding and noncoding variants associated with COPD and COPD-related phenotypes. In addition, WGS of COPD will provide insights into genetic determinants of omics data types, enable studies of resistant smokers in TOPMed, and empower better genotype imputation in studies without WGS data.

Murine Models of COPD Susceptibility Genes

Although GWAS have identified multiple genomic regions that contain COPD susceptibility determinants, the key gene (or genes) in those regions is typically unknown; multiple genes can be located near GWAS signals, and functional variants often do not influence the nearest gene. One approach to

implicate the key gene within a COPD GWAS region is to determine whether orthogonal evidence exists to support a particular gene in COPD pathogenesis from animal models. As shown in Table 1, there are currently seven genes that are located within (or near) genome-wide significant association regions for COPD and that have been implicated in emphysema pathogenesis on the basis of a knock-out or transgenic murine model. These seven genes (*HHIP* [9], *FAM13A* [10], *IREB2* [11], *AGER* [12], *MMP1* [13], *MMP12* [14], and *SFTPD* [15]) point to potentially relevant biological processes involved in COPD pathogenesis. Some of these biological processes, such as protease–antiprotease balance (related to *MMP1* and *MMP12*), have been known for decades; however, other biological processes, such as mitochondrial iron (related to *IREB2*) have not been widely studied before the GWAS era.

Omics Approaches to Identify Key Genes and Functional Variants within COPD GWAS Loci

As shown in Figure 1, GWAS have been extremely successful in identifying genomic regions that are clearly statistically associated with complex diseases and other phenotypes; more than 60,000 genome-wide significant SNP–phenotype associations have been reported. However, the functional variants responsible for those associations have only been identified in a tiny fraction of associated regions. Integration of GWAS results with omics data, including transcriptomics, epigenetics, proteomics, and metabolomics, can help to identify the key gene and the functional variants driving those GWAS associations.

Table 1. Functional validation of chronic obstructive pulmonary disease genome-wide association study genes in murine models

Gene	Reference	Model	Phenotype	Postulated Biological Effect/Pathway
<i>MMP1</i> <i>MMP12</i> <i>SFTPD</i>	D'Armiento <i>et al.</i> , 1992 (13) Hautamaki <i>et al.</i> , 1997 (14) Wert <i>et al.</i> , 2000 (15)	Transgenic Knock-out Knock-out	Increased emphysema Decreased emphysema Increased emphysema	Collagenase activity Metalloelastase activity Matrix metalloproteinase activity
<i>HHIP</i> <i>AGER</i> <i>IREB2</i>	Lao <i>et al.</i> , 2015 (9) Sambamurthy <i>et al.</i> , 2015 (12) Cloonan <i>et al.</i> , 2016 (11)	Heterozygous knock-out Knock-out Knock-out	Increased emphysema Decreased emphysema Decreased emphysema and airway disease	Lymphocyte activation Neutrophil recruitment Mitochondrial iron
<i>FAM13A</i>	Jiang <i>et al.</i> , 2016 (10)	Knock-out	Decreased emphysema	Wnt/ β catenin

Figure 1. Disconnect between GWAS associations and functional variant identification in complex diseases and related phenotypes. GWAS = genome-wide association study. Notes: 1) For more details, see Reference 21. 2) GWAS Associations from GRASP at <https://grasp.nhlbi.nih.gov/Search.aspx> on June 3, 2018.

Lamontagne and colleagues integrated expression quantitative trait locus (eQTL) data from 1,038 lung tissue samples with International COPD Genetics Consortium GWAS results (16). They assessed transcriptome-wide eQTLs, Bayesian colocalization, and Mendelian randomization evidence for shared effects of association signals on eQTL and COPD GWAS. They identified the statistically most likely causal gene in 32% of loci that included multiple genes, including *DSP*. However, not every GWAS SNP will be an eQTL in lung tissue samples.

The COPD GWAS region on chromosome 15q25 has been intriguing but challenging to understand. That genomic region clearly contains genetic determinants related to smoking behavior and nicotine addiction (17), but it also appears to contain COPD genetic determinants unrelated to nicotine addiction (18)—potentially explained by *IREB2* (11). Nedeljkovic and colleagues recently studied four COPD GWAS SNPs on chromosome 15q25 in the Rotterdam study (19). They examined eQTLs in 1,087 lung tissue samples and methylation QTLs in 1,489 blood samples. They found that all four SNPs were eQTLs and methylation QTLs for *IREB2* (as well as for other genes in that region), and they also found differential methylation in COPD cases versus control subjects for CpG sites near *IREB2*. They concluded that DNA methylation may affect *IREB2* expression to influence COPD risk; however, the functional variants for COPD pathogenesis on 15q25 are still unclear. Of interest, a trans-eQTL was also found for *FAM13A* in this 15q25 region, potentially suggesting a

biological network connection between COPD GWAS loci on chromosomes 15q25 and 4q22.

Efforts to identify functional variants systematically within GWAS regions have traditionally been laborious and time consuming. Zhou and colleagues identified a likely functional variant in the *HHIP* COPD GWAS region by using chromatin conformation capture (3C) assays that demonstrated a localized segment of the GWAS region interacted with the *HHIP* promoter (20). Subsequently, they demonstrated enhancer activity within a 500-bp region and an SNP that altered binding to a transcription factor (SP3) that likely regulates *HHIP* (20). More recent efforts using higher-throughput approaches like massively parallel reporter assays hold promise for increasing the rate of functional variant identification in complex diseases (21).

Network-based Approaches to COPD Susceptibility and Heterogeneity

With a few exceptions (e.g., *AGER* [22]), COPD GWAS genetic determinants are primarily located in noncoding, regulatory regions of the genome. Although likely COPD susceptibility genes typically do not show substantial differences between COPD cases and control subjects for lung tissue gene expression levels, Morrow and colleagues noted that the genes that they interact with (based on protein–protein interactions, trans-eQTLs, etc.) are often differentially expressed

in relation to COPD disease status or lung function (23). Thus, COPD susceptibility is determined by a network of interacting genes and proteins. Boyle and colleagues speculated that there are primary genetic drivers of a complex disease, and genes that are connected to these primary drivers will also influence disease susceptibility through their network connections to the primary drivers (24). This omnigenic model is concordant with network medicine views of complex diseases (25).

Multiple approaches can be used to identify the network components of a complex disease like COPD, including correlation-based networks, gene regulatory networks, protein–protein interaction networks, and Bayesian networks. Morrow and colleagues used weighted gene coexpression networks, a form of correlation-based networks, to identify a disease network module that was differentially expressed between COPD cases and controls; this network module was strongly related to B lymphocyte function (23). Network approaches can also be used to investigate COPD heterogeneity; Chang and colleagues applied network-based stratification to gene expression data, and they identified clusters of subjects with COPD on the basis of gene expression that were also related to clinical characteristics of COPD heterogeneity in both the ECLIPSE (Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points) and COPDGene (Genetic Epidemiology of COPD) study populations (26). Further research will be required to identify cell- and tissue-specific networks and to validate network relationships in the laboratory.

Unresolved Questions and Future Research Directions

Although substantial progress has been made in COPD genetics over the past 3 decades, many unresolved issues remain to be addressed. Only small percentages of the estimated genetic contributions to COPD susceptibility (27) and lung function levels (4) can be explained by known genetic loci. Larger genetic association studies will be required to create a more comprehensive set of COPD risk loci. The role of rare genetic variants in COPD genetic risk and heterogeneity should be investigated with whole-exome sequencing and/or WGS studies, because

the relative impact of common versus rare genetic variants in COPD is still uncertain. It is unclear whether all genetic determinants of COPD will be primarily genetic determinants of lung function in general population samples or if a subset will be related directly to lung parenchymal and airway destruction without influencing spirometric values in nonsmoking general population samples. In addition to larger genetic association studies, the functional genetic variants in COPD GWAS loci, and the genes that they influence, need

to be identified. The relationships between COPD GWAS genes need to be characterized, and the impact of the biological networks related to COPD susceptibility genes on COPD pathogenesis, heterogeneity, and progression needs to be understood. The epigenetic marks that influence COPD risk should be found, and the interactions between epigenetic and genetic risk factors should be interrogated. Finally, we need to determine how to integrate different omics data types most effectively to provide insights into

COPD pathogenesis and heterogeneity. Progress in these areas could lead to improved understanding of COPD pathogenesis that will likely lead to more accurate diagnosis and prognosis as well as molecular targets for new treatment development. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

Acknowledgment: The author thanks Drs. Michael Cho, Dawn DeMeo, and Craig Hersh for helpful comments regarding this manuscript.

References

- Silverman EK, Chapman HA, Drazen JM, Weiss ST, Rosner B, Campbell EJ, et al. Genetic epidemiology of severe, early-onset chronic obstructive pulmonary disease: risk to relatives for airflow obstruction and chronic bronchitis. *Am J Respir Crit Care Med* 1998;157:1770–1778.
- Hardin M, Silverman EK. Chronic obstructive pulmonary disease genetics: a review of the past and a look into the future. *Chronic Obstr Pulm Dis (Miami)* 2014;1:33–46.
- Hobbs BD, de Jong K, Lamontagne M, Bosse Y, Shrine N, Artigas MS, et al., COPDGene Investigators, ECLIPSE Investigators, LifeLines Investigators, SPIROMICS Research Group, International COPD Genetics Network Investigators, UK BiLEVE Investigators, International COPD Genetics Consortium. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat Genet* 2017;49:426–432.
- Wain LV, Shrine N, Artigas MS, Erzurumluoglu AM, Noyvert B, Bossini-Castillo L, et al. Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. *Nat Genet* 2017;49:416–425.
- Busch R, Hobbs BD, Zhou J, Castaldi PJ, McGeachie MJ, Hardin ME, et al.; National Emphysema Treatment Trial Genetics; Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-Points; International COPD Genetics Network; COPDGene Investigators. Genetic association and risk scores in a chronic obstructive pulmonary disease meta-analysis of 16,707 subjects. *Am J Respir Cell Mol Biol* 2017;57:35–46.
- Stanley SE, Chen JJ, Podlevsky JD, Alder JK, Hansel NN, Mathias RA, et al. Telomerase mutations in smokers with severe emphysema. *J Clin Invest* 2015;125:563–570.
- Qiao D, Lange C, Beaty TH, Crapo JD, Barnes KC, Bamshad M, et al.; Lung GO; NHLBI Exome Sequencing Project; COPDGene Investigators. Exome sequencing analysis in severe, early-onset chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2016;193:1353–1363.
- Schaid DJ, Chen W, Larson NB. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet* 2018;19:491–504.
- Lao T, Glass K, Qiu W, Polverino F, Gupta K, Morrow J, et al. Haploinsufficiency of Hedgehog interacting protein causes increased emphysema induced by cigarette smoke through network rewiring. *Genome Med* 2015;7:12.
- Jiang Z, Lao T, Qiu W, Polverino F, Gupta K, Guo F, et al. A chronic obstructive pulmonary disease susceptibility gene, FAM13A, regulates protein stability of β -catenin. *Am J Respir Crit Care Med* 2016;194:185–197.
- Cloonan SM, Glass K, Laicho-Contreras M, Bhashyam AR, Cervo M, Pabon MA, et al. Mitochondrial iron chelation ameliorates cigarette smoke-induced bronchitis and emphysema in mice. *Nat Med* 2016;22:163–174.
- Sambamurthy N, Leme AS, Oury TD, Shapiro SD. The receptor for advanced glycation end products (RAGE) contributes to the progression of emphysema in mice. *PLoS One* 2015;10:e0118979.
- D'Armiento J, Dalal SS, Okada Y, Berg RA, Chada K. Collagenase expression in the lungs of transgenic mice causes pulmonary emphysema. *Cell* 1992;71:955–961.
- Hautamaki RD, Kobayashi DK, Senior RM, Shapiro SD. Requirement for macrophage elastase for cigarette smoke-induced emphysema in mice. *Science* 1997;277:2002–2004.
- Wert SE, Yoshida M, LeVine AM, Ikegami M, Jones T, Ross GF, et al. Increased metalloproteinase activity, oxidant production, and emphysema in surfactant protein D gene-inactivated mice. *Proc Natl Acad Sci USA* 2000;97:5972–5977.
- Lamontagne M, Berube JC, Obeidat M, Cho MH, Hobbs BD, Sakornsakolpat P, et al. Leveraging lung tissue transcriptome to uncover candidate causal genes in COPD genetic associations. *Hum Mol Genet* 2018;27:1819–1829.
- Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet* 2010;42:441–447.
- Siedlinski M, Tingley D, Lipman PJ, Cho MH, Litonjua AA, Sparrow D, et al.; COPDGene and ECLIPSE Investigators. Dissecting direct and indirect genetic effects on chronic obstructive pulmonary disease (COPD) susceptibility. *Hum Genet* 2013;132:431–441.
- Nedeljkovic I, Carnero-Montoro E, Lahousse L, van der Plaat DA, de Jong K, Vonk JM, et al. Understanding the role of the chromosome 15q25.1 in COPD through epigenetics and transcriptomics. *Eur J Hum Genet* 2018;26:709–722.
- Zhou X, Baron RM, Hardin M, Cho MH, Zielinski J, Hawrylkiewicz I, et al. Identification of a chronic obstructive pulmonary disease genetic determinant that regulates HHIP. *Hum Mol Genet* 2012;21:1325–1335.
- Musunuru K, Bernstein D, Cole FS, Khokha MK, Lee FS, Lin S, et al. Functional assays to screen and dissect genomic hits: doubling down on the national investment in genomic research. *Circ Genom Precis Med* 2018;11:e002178.
- Park SJ, Kleffmann T, Hessian PA. The G82S polymorphism promotes glycosylation of the receptor for advanced glycation end products (RAGE) at asparagine 81: comparison of wild-type RAGE with the G82S polymorphic variant. *J Biol Chem* 2011;286:21384–21392.
- Morrow JD, Zhou X, Lao T, Jiang Z, DeMeo DL, Cho MH, et al. Functional interactors of three genome-wide association study genes are differentially expressed in severe chronic obstructive pulmonary disease lung tissue. *Sci Rep* 2017;7:44232.
- Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic. *Cell* 2017;169:1177–1186.
- Loscalzo J, Barabási AL, Silverman EK, editors. Network medicine: complex systems in human disease and therapeutics. Cambridge, MA: Harvard University Press; 2017.
- Chang Y, Glass K, Liu YY, Silverman EK, Crapo JD, Tal-Singer R, et al. COPD subtypes identified by network-based clustering of blood gene expression. *Genomics* 2016;107:51–58.
- Zhou JJ, Cho MH, Castaldi PJ, Hersh CP, Silverman EK, Laird NM. Heritability of chronic obstructive pulmonary disease and related phenotypes in smokers. *Am J Respir Crit Care Med* 2013;188:941–947.